

One Size Does Not Fit All: A Pluralistic Approach to Mediator Performance Testing and Quality Assurance

ROBERT A. BARUCH BUSH*

I. INTRODUCTION: QUALITY ASSURANCE AND PERFORMANCE TESTING

The mediation field has faced many challenges in the past three decades, as the use of mediation has grown more and more widespread. One key challenge has been the difficulty of assuring that mediators practice in a manner that is both ethical and competent. In part, the difficulty stems from the fact that mediations are generally conducted privately, confidentially, and without any record of the proceeding. As a result, there is limited potential for effective oversight of mediator practices, especially since many parties to mediation lack sophistication about what to expect of mediators and therefore when to complain about their conduct. In short, with an informal, private, and largely unsupervised process like mediation, the potential for poor quality practice is significant and the power to directly police it in specific cases is limited.¹

The response to this challenge has included a whole array of “quality assurance” strategies. A recent report to one statewide, multi-agency committee surveyed quality assurance strategies in use across the country and found that they include, among others: required paper credentials (documented educational or professional background, etc.), required training in an approved mediator training program, mentoring or supervision by

* Rains Distinguished Professor of ADR Law, Hofstra University School of Law. President and Fellow, Institute for the Study of Conflict Transformation, Inc. This Article is based on a presentation made as part of a Symposium on “Assuring Mediator Quality,” held in 2002 at the University of Maryland School of Law and cosponsored by the Institute for the Study of Conflict Transformation at Hofstra Law School, the Center for Dispute Resolution at U.Md. Law School, and the Maryland Judiciary’s Mediation and Conflict Resolution Office.

¹ See SARAH R. COLE ET AL., *MEDIATION: LAW, POLICY, PRACTICE* § 2:8 (2d. ed. 2001); Robert A. Baruch Bush, *Efficiency and Protection, or Empowerment and Recognition?: The Mediator’s Role and Ethical Standards in Mediation*, 41 U. FLA. L. REV. 253, 254–55 (1989); JAY FOLBERG & ALISON TAYLOR, *MEDIATION: A COMPREHENSIVE GUIDE TO RESOLVING CONFLICTS WITHOUT LITIGATION* 244 (1984); Richard Delgado et al., *Fairness and Formality: Minimizing the the Risk of Prejudice in Alternative Dispute Resolution*, 1985 WIS. L. REV. 1359, 1374, 1387–89; Trina Grillo, *The Mediation Alternative: Process Dangers for Women*, 100 YALE L.J. 1545, 1585–90 (1991).

experienced practitioners, written tests, monitoring and user evaluations, complaint or grievance procedures, and performance testing.² The last of these strategies, performance testing, is the focus of this Article, both because it offers great potential to assure quality practice and also because it highlights some of the most controversial issues involved in quality assurance efforts. Performance testing is done both in live cases and through videotaped simulations. In either method, the mediator is required to demonstrate his or her skills in the context of an unfolding mediation situation, and competency is assessed by an evaluator according to some preestablished set of relatively detailed behavioral measures.³

Performance testing, as a quality assurance strategy, has presented a paradox of sorts. On the one hand, it has been viewed as one of the best ways to measure and guarantee the quality of practice, as when the Commission on Qualifications of the Society for Professionals in Dispute Resolution (hereinafter, SPIDR Commission) recommended in 1989 that qualifications for practice "should be based on performance, emphasizing the knowledge and particular skills necessary for competent practice. . . . Policymakers should adopt . . . performance criteria . . . and incorporate performance-based testing into training and apprenticeship programs."⁴ On the other hand, relatively few jurisdictions, agencies, or programs have actually adopted performance testing,⁵ and even SPIDR's successor organization, the Association for Conflict Resolution (ACR) has been advised in a task force report to reject performance-testing as part of its process for certifying mediators.⁶

² CHARLES POU, JR., *MEDIATOR QUALITY ASSURANCE: A REPORT TO THE MARYLAND MEDIATOR QUALITY ASSURANCE OVERSIGHT COMMITTEE* 8-14 (Feb. 2002) (on file with author and available from Maryland Mediation and Conflict Resolution Office).

³ *Id.*; see also TEST DESIGN PROJECT, *PERFORMANCE-BASED ASSESSMENT: A METHODOLOGY, FOR USE IN SELECTING, TRAINING AND EVALUATING MEDIATORS* 7-21 (1995) [hereinafter *METHODOLOGY*].

⁴ SOCIETY OF PROFESSIONALS IN DISPUTE RESOLUTION COMMISSION ON QUALIFICATIONS, *QUALIFYING NEUTRALS: THE BASIC PRINCIPLES* 4 (1989), reprinted in JAMES J. ALFINI ET AL., *MEDIATION THEORY AND PRACTICE* 344-46 (2001).

⁵ See Ellen Waldman, *Credentialing Approaches: The Slow Movement Toward Skills-Based Testing Continues*, *DISP. RESOL. MAG.*, Fall 2001, at 13, 16, available at www.convenor.com/madison/waldman.htm; COLE ET AL., *supra* note 1, § 11.2; POU, *supra* note 2, at 23-24.

⁶ In its report to the ACR Board of Directors, the Task Force on Mediator Certification recommended a process that uses a "written knowledge assessment" rather than a performance test in order to gauge mediator competence, because of questions about whether performance testing is "reliable," "valid," "practical," and financially feasible. See ACR MEDIATOR CERTIFICATION TASK FORCE REPORT AND

There are several reasons for the ambivalence about performance testing. First, performance tests cannot be constructed without defining the specific skills a competent mediator should possess and demonstrate; and despite a considerable effort to do this, the task of defining the required skills repertoire has proved difficult.⁷ Second, though not unrelated, performance testing necessarily means exposure to judgments about competency and potential exclusion from practice, and mediators in general have not yet been willing to subject themselves to such consequences.⁸ Finally, performance testing is likely to be costly in both time and money, and a field that claims a cost advantage over other processes is understandably resistant to front-end costs that may drive up the price of mediators' services.⁹

Despite these and other barriers, the push for performance testing has continued and grown, for another set of compelling reasons. One important reason is that, as the use of mediation has expanded and gained institutional support, institutional users have sometimes demanded that providers have an objective, justifiable basis for recommending particular mediators.¹⁰ Without such a basis, there is potential for claims of unfairness and discrimination by other mediators, and no one wants to defend against such claims. Perhaps even more important, mediators themselves recognize that, with the process being used more and more widely, regulation will eventually be imposed on the field if it does not regulate quality for itself.¹¹ Indeed, in California, a state with one of the longest histories of using mediation as a court-annexed process, legislation was introduced to require performance testing.¹² While the legislation did not pass, the effort was indicative of the potential for

RECOMMENDATIONS TO THE ACR BOARD OF DIRECTORS (March 31, 2004) [hereinafter ACR TASK FORCE], available at www.acrnet.org/about/taskforces/certification.htm.

⁷ See Waldman, *supra* note 5, at 15–16; Paul J. Spiegelman, *Certifying Mediators: Using Selection Criteria to Include the Qualified—Lessons from the San Diego Experience*, 30 U.S.F. L. REV. 677, 682 (1996).

⁸ See Linda C. Neilson & Peggy English, *The Role of Interest-Based Facilitation in Designing Accreditation Standards: The Canadian Experience*, 18 MEDIATION Q. 221, 223 (2001); Barbara Filner & Michael Jenkins, *Performance-Based Evaluation of Mediators: The San Diego Mediation Center's Experience*, 30 U.S.F. L. REV. 647, 654 (1996); METHODOLOGY, *supra* note 3, at 7–8.

⁹ See Waldman, *supra* note 5, at 14; ACR TASK FORCE, *supra* note 6; METHODOLOGY, *supra* note 3, at 7–8.

¹⁰ See, e.g., Neilson & English, *supra* note 8, at 227.

¹¹ See, e.g., Filner & Jenkins, *supra* note 8, at 654.

¹² See Donald T. Weckstein, *Mediator Certification: Why and How*, 30 U.S.F. L. REV. 757, 757–59, 793–801. The appendix to this article includes Recommended Statute based on California SB 1428, the failed legislation referred to in the text. *Id.* at app.

outside regulation, and this potential has itself continued to fuel the discussion of performance testing as a self-regulatory measure.

The thesis of this Article is that efforts to generate support for performance testing, and to develop valid and accepted performance tests, have been undermined by a common and serious flaw, which has resulted in both the difficulty of constructing such tests and the resistance to adopting them. The flaw is the underlying premise that there is a single set of “core skills” that any mediator must possess and demonstrate in order to be considered a competent practitioner. This premise is related to an even deeper one—that differences in mediator practice are matters of style rather than principle, and that different “models” of mediation are not fundamentally distinct from one another. The corollary of these premises is that no matter what “model” of mediation a practitioner employs, she or he should possess the same basic set of skills and therefore can be evaluated by the same performance test. If this premise of homogeneity—or “mediation as monolith”—were abandoned, performance testing would become both more practical and more acceptable, and the effort to assure the quality of practice would be greatly enhanced.

In this Article, this thesis is developed in several steps. Part II offers a brief history of the earliest significant efforts to develop mediator performance testing, showing how those efforts were grounded on the “core skills” and “homogeneity” premises. Part III then presents several examples of performance tests that followed the example of that early model. Part IV discusses the development in the field of the recognition that there are different and distinct models of mediation in use and traces the impact of that recognition on performance testing, analyzing two specific examples of performance tests, showing how each addresses the issue of “models” and concluding that although there has been significant improvement in this area, no current testing regime is sufficiently pluralistic in its approach to different models of practice. Finally, Part V suggests what a pluralistic approach to performance testing might look like and argues that such an approach is both practical and desirable.

II. PERFORMANCE TESTING AND “THE COMMON CORE” PREMISE: THE TEST DESIGN PROJECT

The best-known early effort to establish a method for performance testing began roughly fifteen years ago, in the late 1980s. Launched in response to the abovementioned SPIDR Commission’s recommendation that standards be performance-based, the “Test Design Project” (hereinafter TDP) was an independent, grant-funded “effort to provide mediation programs, courts and other interested parties with improved tools for selecting, training

and evaluating mediators," especially through the development of performance testing.¹³ The TDP published its first report, "Interim Guidelines for Selecting Mediators," in 1993, including a "model performance test,"¹⁴ which generated a number of critiques.¹⁵ This led to further work culminating in the publication in 1995 of a final report describing a "Methodology" for performance-based assessment, rather than a single "model test."¹⁶ Nevertheless, the TDP's Methodology did contain a number of model tests, including the original from the Interim Guidelines, and that original test (hereinafter, TDP Test) remained very influential in subsequent efforts by others to design and implement performance testing, as will be discussed below.¹⁷

A. *The Context of the TDP Test: "Common Core" Skills and Mediation Theory*

Both the specific content and the underlying premises of the TDP Test are significant for the analysis here. Regarding content, the TDP Test focused on the following six categories of mediator activity as the focus for the evaluation of performance: investigating to gather information, conveying empathy, projecting impartiality, generating options, generating agreements, and managing the interaction. For each category, specific behaviors were identified and grouped so as to distinguish between high, medium, and low levels of competency (scaled as 3, 2, or 1 point behaviors).¹⁸ The test was to be applied by trained evaluators based on observation of a simulated mediation.¹⁹

¹³ METHODOLOGY, *supra* note 3, at 1. The quoted language is taken from the final report of the TDP, published two years after the publication of its first test. *See infra* text accompanying notes 35–38.

¹⁴ TEST DESIGN PROJECT, INTERIM GUIDELINES FOR SELECTING MEDIATORS 7–10 (1993), *reprinted* in Christopher Honeyman, *A Consensus on Mediators' Qualifications*, 9 NEGOT. J. 295, 302–05 (1993) [hereinafter Honeyman].

¹⁵ *See, e.g.*, Richard A. Salem, *The "Interim Guidelines" Need a Broader Perspective*, 9 NEGOT. J. 309 (1993); Craig A. McEwen, *Competence and Quality*, 9 NEGOT. J. 317 (1993); Carrie Menkel-Meadow, *Measuring Both the Art and Science of Mediation*, 9 NEGOT. J. 321 (1993); Robert A. Baruch Bush, *Mixed Messages in the "Interim Guidelines"*, 9 NEGOT. J. 341 (1993).

¹⁶ METHODOLOGY, *supra* note 3.

¹⁷ *See infra* text accompanying notes 39–59.

¹⁸ Honeyman, *supra* note 14, at 302–05.

¹⁹ *See* METHODOLOGY, *supra* note 3, at 34–38. *See generally* David E. Matz, *Some Advice for Mediator Evaluators*, 9 NEGOT. J. 327 (1993).

Before analyzing further the specifics of the test, it is important to place it in context. First, its overall message is that mediators display competency when they are effective in organizing and focusing the discussion of issues, and then moving parties from disagreement to agreement on those issues and producing a concrete settlement. There are actually two premises embedded in this message: first, that there is a "common core" of behaviors involved in the work of effective mediators; and second, that these behaviors are the means to an end that comprises the ultimate goal of mediation—achievement of an agreement that settles the parties' dispute. These premises are quite explicit in the work that formed both the immediate and the larger background for the development of the TDP Test.

The immediate background was the work of Christopher Honeyman, the TDP's director and draftsman, who had been working since the mid-1980s to identify a "common core" of behaviors exhibited by effective mediators.²⁰ In Honeyman's work, it is clear that the meaning of "effectiveness" is success in attaining a settlement. Each of the elements he finds "common" to the work of the mediators he studied—investigation, empathy, persuasion, invention, and distraction—are described in terms of their usefulness in promoting settlement, which is taken for granted as the goal of their work.²¹ In building a performance test on the foundation of this work, the TDP was thus developing a test that would gauge a mediator's ability to achieve the goal of settlement production.

As for the larger background, at the time of Honeyman's initial work and continuing through the TDP's development of its original model test, this conception of the goal of mediation, and the resulting view of how the process is most effectively conducted, was essentially unchallenged. That is, there was only one model of mediation known or imagined, and it was a model focused on the production of agreements.²² Although differences in "styles" of practice had been identified and studied by researchers, including such distinctions as "orchestrators versus dealmakers"²³ or "bargaining versus therapeutic,"²⁴ all of these were regarded as stylistic variations on the

²⁰ See, e.g., Christopher Honeyman, *Five Elements of Mediation*, 4 NEGOT. J. 149 (1988) [hereinafter *Five Elements*]; Christopher Honeyman, *On Evaluating Mediators*, 6 NEGOT. J. 23 (1990); Christopher Honeyman, *The Common Core of Mediation*, 8 MEDIATION Q. 73 (1990).

²¹ See METHODOLOGY, *supra* note 3, at 15–16; *Five Elements*, *supra* note 20, at 153–55.

²² See ROBERT A. BARUCH BUSH & JOSEPH P. FOLGER, *THE PROMISE OF MEDIATION: RESPONDING TO CONFLICT THROUGH EMPOWERMENT AND RECOGNITION* 55–68 (1994).

²³ See DEBORAH M. KOLB, *THE MEDIATORS* 23–45 (1983).

²⁴ See Susan S. Silbey & Sally E. Merry, *Mediator Settlement Strategies*, 8 LAW & POL'Y 7, 19–25 (1986).

common theme of how best to produce settlements.²⁵ It was only in the mid-1990s that it was first argued that mediators differed not merely in the way they approached the common goal of settlement, but also—and more profoundly—in the very conception of what goal they were pursuing, and that some mediators were following a genuinely different model of practice aimed at a goal other than settlement.²⁶ Thus, the assumption of homogeneity that informed the TDP was consistent with the field's common understanding of the mediation process at that time.

Moreover, apart from a common view of mediation's goal, the field in the 1980s and early 1990s also had a fairly consistent view of how mediators could achieve that goal in practice. Thus, there were numerous, fairly detailed "guides to practice" published prior to Honeyman's work, including classic books and articles by authors still recognized as authorities today, such as Moore,²⁷ Stulberg,²⁸ Folberg and Taylor,²⁹ Saposnek,³⁰ Haynes,³¹ and others. The "theory and practice of mediation" described in that literature—and still followed by many today—involves the mediator leading the parties through a sequence of stages: opening the session and setting ground rules, gathering information, defining issues, generating options, generating movement (by persuasion), and achieving agreement and closure.³² The description of stages and strategies differ from text to text, but the commonalities are very clear, as to both the goal and the means to achieve it. What is also clear is the principle that in all these stages of mediation, the mediator is the one who controls and conducts the process at every stage, and

²⁵ See BUSH & FOLGER, *supra* note 22, at 59–63; Silbey & Merry, *supra* note 24, at 19–25.

²⁶ See BUSH & FOLGER, *supra* note 22, at 81–95.

²⁷ CHRISTOPHER W. MOORE, *THE MEDIATION PROCESS: PRACTICAL STRATEGIES FOR RESOLVING CONFLICT* (1986).

²⁸ Joseph B. Stulberg, *The Theory and Practice of Mediation: A Reply to Professor Susskind*, 6 VT. L. REV. 85 (1981); JOSEPH B. STULBERG, *TAKING CHARGE/MANAGING CONFLICT* (1987).

²⁹ FOLBERG & TAYLOR, *supra* note 1.

³⁰ DONALD SAPOSNEK, *MEDIATING CHILD CUSTODY DISPUTES: A SYSTEMATIC GUIDE FOR FAMILY THERAPISTS, COURT COUNSELORS, ATTORNEYS, AND JUDGES* (1983).

³¹ JOHN M. HAYNES, *DIVORCE MEDIATION* (1981); JOHN M. HAYNES & GRETCHEN L. HAYNES, *MEDIATING DIVORCE: CASEBOOK OF STRATEGIES FOR SUCCESSFUL FAMILY NEGOTIATIONS* (1989).

³² See ALFINI ET AL., *supra* note 4, at 107–40. The Alfini "casebook" offers a good summary based on the work of the above-cited mediation authors and others. See also Dorothy J. Della Noce, *Mediation as a Transformative Process: Insights on Structure and Movement*, in *DESIGNING MEDIATION: APPROACHES TO TRAINING AND PRACTICE WITHIN A TRANSFORMATIVE FRAMEWORK* 71, 71–76, 72 n.1 (J.P. Folger & R.A.B. Bush eds., 2001) [hereinafter *DESIGNING MEDIATION*].

effective mediation practice requires the exercise of considerable control, direction, and influence to “keep the process moving” toward the goal of settlement.³³ This operating principle of mediator process control, despite the centrality of the value of self-determination in the mediation process, is often explained with the conventional wisdom that “the parties control the outcome, but the mediator controls the process.”³⁴

B. The TDP Test: Defining and Measuring Competence Behaviorally

Within this larger context of an authoritative literature describing an accepted view of the goal of mediation and the practices needed to achieve it, the specifics of the TDP Test make a great deal of sense. In four of the categories of mediator activity measured by the TDP Test, the high-scoring behaviors included the following:

- *Investigation*: . . . identifying and seeking out relevant information pertinent to the case.
 . . . Detected and addressed hidden issues . . . Defined and clarified the issues . . . Gathered information through incisive, tough and uncomfortable questions.
- *Generating Options*: . . . generation of ideas and proposals . . .
 Generated, assessed and prioritized alternative solutions . . . Recognized underlying problems as opposed to symptoms . . . Invented and recommended unusual but workable solutions consistent with case facts. Vigorously pursued avenues of collaboration between the parties.

³³ See, e.g., Stulberg, *supra* note 28, at 97–106; STULBERG, *supra* note 28, at 95–106; HAYNES & HAYNES, *supra* note 31, at 3, 16–17; see also Della Noce, *supra* note 32, at 74; Deborah M. Kolb & Kenneth Kressel, *The Realities of Making Talk Work*, in WHEN TALK WORKS: PROFILES OF MEDIATORS 459, 470–74 (Deborah M. Kolb & Associates eds., 1994).

³⁴ See, e.g., John M. Haynes, *Mediation and Therapy: An Alternative View*, 10 MEDIATION Q. 21, 23–24 (1992); HAYNES & HAYNES, *supra* note 31, at 16; Stulberg, *supra* note 28, at 96. But see Joseph P. Folger, *Who Owns What in Mediation?: Seeing the Link Between Process and Content*, in DESIGNING MEDIATION, *supra* note 32, at 55 (challenging the distinction between process and content control).

- *Generating Agreements*: . . . moving the parties toward finality and “closing” an agreement.

. . . Emphasized areas of agreement. Clarified and framed points of agreement . . . Asked tough questions to highlight unreasonable positions . . . Packaged and linked issues to demonstrate mutual gains from agreements.

- *Managing the Interaction*: . . . developing strategy, managing the process . . .

Had effective techniques for redirecting parties’ focus away from sullen or otherwise unproductive colloquies. Maintained optimism and forward movement, emphasized progress, showed tenacity . . . Made all decisions about caucusing, order of presentation, etc., consistent with rationale for progress toward resolution . . . ³⁵

Specifying the above behaviors as indicators of high competence—whether taken together or separately—clearly reflects the premise that the goal of mediation is a settlement agreement, the production of which defines success and demonstrates effective and competent practice. It also reflects the view that, in order to achieve this goal and be effective, a mediator must engage in behaviors that are, in varying degrees, controlling, directive, forceful, and manipulative. In short, the TDP Test’s framework for evaluating competency closely reflects the view prevalent in the field at that time regarding the goal of mediation and the strategies needed to achieve it. Put differently, the TDP Test is reflective of an approach to mediation, a “model” of mediation, which was seen at the time as the single, universally accepted meaning of mediation in general. Viewed in this light, the TDP Test represented a significant contribution, because it went a long way toward operationalizing that model of mediation in a way that would allow its practitioners’ competency to be performance-tested.

Despite this value, the TDP Test evoked significant criticism. One type of critique pointed to errors or omissions of specific behaviors important to effective settlement-oriented mediation.³⁶ Another and more serious type of critique questioned the general validity of the test, suggesting that, precisely because it was tied so closely to a settlement-oriented view of mediation, the TDP Test was inadequate as an instrument for measuring the competency of mediators practicing other models of mediation.³⁷ Of course, the premise of this critique was the existence of distinct and different models of practice, a

³⁵ Honeyman, *supra* note 14, at 303–05 (bullets added).

³⁶ See, e.g., Salem, *supra* note 15, at 309; Menkel-Meadow, *supra* note 15, at 321.

³⁷ See Bush, *supra* note 15.

notion that gained visibility in the field after the TDP was launched and had a significant impact on subsequent efforts to develop mediator performance tests as a form of quality assurance.³⁸ That “second phase” of performance test work is the subject of part IV below. First, however, it is important to document the continued vitality of the type of test originated by the TDP—a test that primarily measures the competence of mediators by reference to their effectiveness in settlement-producing practices.

III. TESTING FOR EFFECTIVENESS IN PROBLEM-SOLVING AND AGREEMENT-PRODUCTION: PROGENY OF THE TDP TEST

Although the TDP Test does not appear to have been adopted by any program or jurisdiction exactly as it was drafted, it was quite influential on other performance tests, both as to the general orientation of the tests and the specific behaviors measured. In each of the several tests discussed in this Part, it is quite clear that the ultimate indicator of competent performance is the ability to produce an agreement that settles the dispute. In addition, the specific behaviors measured by these tests tend to be similar if not identical, compared to those measured by the other tests and the TDP Test.

A. *The San Diego Test*

The San Diego Mediation Center (SDMC) is a nonprofit agency that started as a community-based program and now also serves the courts and other government agencies, providing court-annexed mediation and even private divorce mediation.³⁹ In 1993, SDMC instituted a mediator performance test as part of a process for “certifying” mediators. While the test does not appear to have been based directly on the TDP Test, it was clearly developed on the same foundation (i.e., Honeyman’s work on the “core skills” needed for effective mediation).⁴⁰ From that foundation, and like the TDP itself, SDMC developed a test to measure “skills that are appropriate and useful in the mediation process, yet not unique to any one model or style of mediation. . . . The test instrument measures eighteen specific behaviors . . . that experienced mediators would agree must be largely present in almost any effective mediation.”⁴¹ It is clear from this account that the premise of homogeneity was accepted in the development of

³⁸ See *infra* text accompanying notes 59–115.

³⁹ See Spiegelman, *supra* note 7, at 698–99.

⁴⁰ See Filner & Jenkins *supra* note 8, at 656.

⁴¹ *Id.* at 658–59.

SDMC's test (i.e., the test assumes a single, generalized model of practice and tries to measure competency on that basis).

From the details of the SDMC Test, it is also clear that it reflects the same premises regarding the ultimate goal of mediation, and the best strategies for achieving that goal, as those of the TDP Test. Thus, the categories of mediator activity measured—and the specific behavioral indicators used to measure them (on a 5-point scale)—include the following:

- *Process Flow*: . . . Is there clear movement toward a resolution?
- *Opening Statement*: . . . Does it adequately cover procedural information, . . . confidentiality, ground rules and expectations? . . .
- *Facilitating Position Statements*: . . . Does mediator ask necessary clarifying questions and enforce the ground rules when needed? . . .
- *Coordinating the Exchange/Conflict Analysis*: Does mediator have the ability to frame issues for discussion . . . ? Does the mediator set out clear agenda of issues . . . ?
- *Managing the Negotiation*: Does mediator facilitate a productive negotiation ? . . . Is discussion future oriented?
- *Organizing Issues*: Is there evidence of a strategy for prioritizing issues and for overcoming impasse? . . .
- *Neutral Language*: Look at mediator's ability to . . . reframe issues in neutral or positive language, and put disputants' demands into context of "interests."
- *Strategic Development*: As needed, does mediator adjust process and guide clients toward productive interaction and resolution? Is there evidence of pre-planning in room set up, focus on future behavior, and appropriate use of caucus to move process forward?⁴²

It is clear that the premise of these behavioral measures, taken together, is that competent performance involves the ability to produce an agreement that resolves the dispute. It is also clear that the specific behaviors seen as measuring this ability involve the mediator "controlling the process" in a variety of ways in order to achieve the desired outcome. Indeed, in an account of the development of the SDMC Test by two individuals connected to the Center, the authors state that 12 of the 18 behaviors measured "relate

⁴² THE SAN DIEGO MEDIATION CENTER CREDENTIALING PERFORMANCE EVALUATION [hereinafter SDMC TEST], reprinted in METHODOLOGY, *supra* note 3, at 27–30 (bullets added); see also Spiegelman, *supra* note 7, at 705 & n.115.

to the mediator's control of the process."⁴³ As with the TDP Test, the SDMC Test defines and measures mediator competency in terms of the goal of agreement production and a set of directive, process-controlling behaviors likely to achieve that goal. In fact, as will be discussed shortly, there is considerable overlap in the process-controlling behaviors specified in each test. First, however, consider some further examples of performance tests that are close to the TDP Test in their approach, although they are used in live observations of mediations, rather than in simulations done for test purposes.

B. The Maine Test

The Court ADR Services (CADRES) office of the Maine Judiciary employs a performance test for mediators who seek acceptance to one of its several rosters.⁴⁴ The test involves having an evaluator complete an "observer's checklist" after observing the candidate conduct a live mediation session. The CADRES Test includes seven categories to be checked for (a numerical rating is not used), with 10–12 specific behaviors under each. The categories are quite similar to those of the TDP: opening statement, relationship with participants, defining issues, communication, generating and testing options, reaching and confirming agreement, and ending the session.⁴⁵ Once again, some examples of the specific behaviors measured show that the CADRES Test shares the same view as the TDP and SDMC tests about the meaning of mediator competence:

- RELATIONSHIP WITH PARTICIPANTS . . .
 - g. allows reasonable venting of emotions . . .
- DEFINING THE ISSUES
 - a. identifies disputed issues . . .
 - c. develops a strategy for the mediation process . . .
 - e. learns about underlying interests . . .
 - g. elicits and emphasizes the positive
 - h. keeps parties focused on relevant issues
 - i. keeps session on track . . .

⁴³ Filner & Jenkins, *supra* note 8, at 660.

⁴⁴ Phone interview with William Galloway, Private Mediator (Oct. 22, 2003).

⁴⁵ Court Alternative Dispute Resolution Service (CADRES) Observer's Checklist for Mediation (on file with author).

- GENERATING AND TESTING OPTIONS
 - a. helps parties find creative solutions to disputes
 - b. uses a variety of methods to help generate possible solutions
 - c. reframes and restates options and proposals
 - d. encourages parties to negotiate
 - e. uses private caucuses as needed to generate ideas . . .
 - j. projects optimism
- REACHING AND CONFIRMING AGREEMENT
 - a. aims for clear, practical, legal agreements . . .
 - d. emphasizes a forward-looking, problem-solving approach . . .⁴⁶

The clear message of the tested behaviors, taken together, is that competence is measured by the ability to produce agreement, which is presumed to rest on the use of a variety of specific “process control” interventions.

C. The Navy and Virginia Tests

A final example of a TDP-type test is a performance test used by two different mediation programs, with only slight variations. Each of the programs—the Department of the Navy (DON) Certified Mediator Program (for mediations involving civilian employees) and the Supreme Court of Virginia Certified Mediator Program (for court-annexed mediation)—employs its test in observations of actual sessions, by “mentors” who are already certified as mediators.⁴⁷ According to the Virginia Test document, the test was adapted from a peer evaluation form developed by the Community Mediation Center in Harrisonburg, Virginia. Some of the categories and specific behaviors evaluated (on a 1–5 scale) are as follows:

- Introduction . . .
 - Established ground rules . . .
- Issue Clarification
 - Asked appropriate questions
 - Identified interests . . . and underlying problems
 - Identified common ground
 - Reframed statements and issues

⁴⁶ *Id.*

⁴⁷ See Waldman, *supra* note 5, at 15.

- Generation of Options
 - Organized and prioritized mediable issues
 - Focused on present and future needs rather than positions
 - Elicited multiple options and explored settlement possibilities
- Resolution/Closure
 - Facilitated negotiation and bargaining . . .
 - Assisted in developing agreement that is balanced, fair, realistic . . .
 - Sufficient effort exerted to assist parties in reaching agreement
- Special Techniques
 - Demonstrated appropriate use of caucus . . .
 - Overcame impasse, resistance or difficult behavior . . .
 - Dealt with power imbalance or control issues . . .⁴⁸

The specifics of the Navy Test are almost identical.⁴⁹ Again, the examples reveal a clear focus on resolution as the goal, and on “process control” interventions as the means to this end, quite similar to the other tests examined above.

D. Common Patterns in the Tests: A Single Picture of Competence

Beyond the fact that the tests reviewed above, as demonstrated, are similar in general character and purpose, they are also similar in listing many of the same specific behaviors as measures of competency. Identifying these specific overlaps will show even more clearly how all these tests assume a single model of mediation, which any competent practitioner must know how to employ. To make this point more salient, however, it is first important to add a bit more background about how mediation strategy was pictured in the practice literature of the 1980s and early 1990s. While Honeyman’s “core elements” of practice captured much of this picture,⁵⁰ various authorities on “good practice” emphasized certain elements in particular.

Among the most important strategies emphasized in the practice literature were:

⁴⁸ Supreme Court of Virginia, Mentee Evaluation Form, *available at* www.courts.state.va.us/drs/forms/ADR-1001-070199.htm (last visited Apr. 8, 2004) [hereinafter Virginia Test].

⁴⁹ Department of the Navy/Civilian Human Resources, Co-Mediation Evaluation Form, *available at* www.adr.navy.mil/adr/mediacert.asp (last visited Apr. 8, 2004) [hereinafter DON Test].

⁵⁰ See *supra* text accompanying notes 20–21.

- focusing parties away from distributive, positional bargaining and toward integrative, needs-and-interests-based bargaining;⁵¹
- emphasizing “common ground” while deemphasizing areas of disagreement;⁵²
- focusing discussion on future commitments and not past grievances;⁵³ and
- limiting strong emotional expression in order to avoid undermining rational discussion.⁵⁴

Some of these “good mediation” practices were actually derived from the burgeoning literature on “principled” or “problem-solving” negotiation,⁵⁵ dating from publication of Fisher and Ury’s classic work on *Getting to Yes* in 1981.⁵⁶ However, many in the mediation field believed that such practices, if used by mediators, would increase mediator effectiveness in producing more agreements—and indeed agreements of better quality—than otherwise attainable.⁵⁷ Therefore, the literature included these elements more and more as part of the authoritative view of effective mediation.

⁵¹ See, e.g., MOORE, *supra* note 27, at 38–39, 187–98, 208–10; STULBERG, *supra* note 28, at 98, 102–03; Alison Taylor, *A General Theory of Divorce Mediation*, in *DIVORCE MEDIATION: THEORY AND PRACTICE* 61, 71–73 (Jay Folberg & Ann Milne eds., 1988).

⁵² See, e.g., STULBERG, *supra* note 28, at 1, 102–03; HAYNES & HAYNES, *supra* note 31, at 35; MOORE, *supra* note 27, at 64.

⁵³ See, e.g., FOLBERG & TAYLOR, *supra* note 1, at 14; SAPOSNEK, *supra* note 30, at 70; STULBERG, *supra* note 28, at 101; HAYNES & HAYNES, *supra* note 31, at 34.

⁵⁴ See, e.g., SAPOSNEK, *supra* note 30, at 176–77; MOORE, *supra* note 27, at 127–32.

⁵⁵ See, e.g., Carrie Menkel-Meadow, *Toward Another View of Legal Negotiation: the Structure of Problem Solving*, 31 U.C.L.A. L. REV. 754 (1984); DAVID A. LAX & JAMES K. SEBENIUS, *THE MANAGER AS NEGOTIATOR: BARGAINING FOR COOPERATION AND COMPETITIVE GAIN* (1986).

⁵⁶ ROGER FISHER & WILLIAM URY, *GETTING TO YES: NEGOTIATING AGREEMENT WITHOUT GIVING IN* (1981).

⁵⁷ See MOORE, *supra* note 27, at 38–39, 71–72; KARL A. SLAIKEU, *WHEN PUSH COMES TO SHOVE: A PRACTICAL GUIDE TO MEDIATING DISPUTES* 5, 151 (1996); Leonard L. Riskin, *Mediator Orientations, Strategies and Techniques*, 12 ALTERNATIVES 111, 111–12 (1994); Lela P. Love, *The Top Ten Reasons Why Mediators Should Not Evaluate*, 24 FLA. ST. U. L. REV. 937, 937–39 (1997); Spiegelman, *supra* note 7, at 694–95; Robert H. Mnookin & Lee Ross, *Introduction*, in *BARRIERS TO CONFLICT RESOLUTION* 3, 22–24 (Kenneth Arrow et al. eds., 1995).

With this background from the practice literature of the period, together with that mentioned earlier,⁵⁸ it is possible to show that the performance tests discussed thus far—including the TDP Test and the four others described—overlap heavily in the measures they specify as indicators of competent practice. Specifically, comparing the indicators from each test cited above, the following patterns emerge:

- All five tests (as noted above) specify mediator behaviors that involve controlling and structuring the unfolding of the process: establishing and enforcing groundrules, defining issues, structuring the agenda, deciding when to caucus, keeping discussions on track, and in general making process decisions “consistent with progress toward resolution.”
- Most of the tests specify mediator behaviors that involve identifying parties’ underlying needs and interests (as opposed to positions), or encouraging interest-based problem-solving (as opposed to positional bargaining) or both.
- Most of the tests specify mediator behaviors that involve emphasizing common ground and areas of agreement (and de-emphasizing disagreement).
- Most of the tests specify mediator behaviors that involve steering the parties to focus their discussions on the future and to disregard the past.
- Most of the tests specify mediator behaviors that, implicitly or explicitly, discourage and limit expression of negative emotions: “sullenness,” “intense emotions,” “unreasonable venting,” “unproductive interaction,” etc.

This comparative summary demonstrates that these five performance tests, despite different formats and some differences in specific content, reflect a largely consistent and common picture of what mediator competence looks like.

Moreover, the common picture embedded in the tests is consistent with the single view of the mediation process that prevailed in the practice literature during the 1980s and early 1990s, according to which there is only one, homogeneous mediation process—with a single goal and an agreed-

⁵⁸ See *supra* text accompanying notes 27–34.

upon set of elements or strategies, despite many variations in individual practitioners' styles. Thus, notwithstanding their own stylistic differences, the five tests discussed above all serve to measure the same thing: the extent to which an individual mediator competently practices that homogeneous process of mediator-controlled problem-solving, in order to achieve the presumed goal of agreement-production.

IV. PERFORMANCE TESTING WITHOUT A "CORE": RECOGNITION OF "MODELS" OF MEDIATION, AND THE IMPACT ON TESTING FOR MEDIATOR COMPETENCE

Until the mid-1990s, differences in the practices of mediators were viewed as matters of stylistic variations in the practice of a single, homogeneous process. This was referred to above as the premise of homogeneity, and it was strongly supported by a practice literature that presented a fairly coherent picture of that process, with its "core elements" agreed upon by all.⁵⁹ Research that documented differences in practice, and suggested "typologies" of different approaches to the mediation process, largely failed to penetrate the world of practice.⁶⁰ Then, in 1994, three scholarly publications appeared that challenged, in varying degrees, the notion that mediation, as practiced, was a single, homogeneous, monolithic process. Over the last decade, that challenge has substantially succeeded, and it is now widely accepted that there are different models of mediation in practice, with coherent distinctions that go beyond individual stylistic differences.⁶¹ The impact on the development of performance testing, as on the rest of the field, has been significant. Before examining that impact, it is useful to summarize the work that brought the premise of homogeneity into question.

⁵⁹ See *supra* text accompanying notes 27–34 and 51–57

⁶⁰ See *supra* notes 23–25 and accompanying text; see also, e.g., Kenneth Kressel et al., *The Settlement-Oriented vs. the Problem-Solving Style in Custody Mediation*, 50 J. SOC. ISSUES 67 (1994); Kenneth Kressel & Dean G. Pruitt, *Conclusion: A Research Perspective on the Mediation of Social Conflict*, in *MEDIATION RESEARCH* 394, 423–25 (Kenneth Kressel et al. eds., 1989).

⁶¹ See, e.g., ALFINI ET AL., *supra* note 4, at 107, 140–47; Neilson & English, *supra* note 8, at 223–24.

A. Distinct Models of Practice: Three Views and a Common Conclusion

In a brief article intended to clarify the spectrum of mediator techniques, rather than articulate distinct models, Professor Leonard Riskin suggested a distinction between “facilitative” and “evaluative” approaches to mediation practice.⁶² The primary difference is the extent to which the mediator evaluates the parties’ substantive proposals and arguments, and recommends terms of settlement, rather than simply managing the process itself. This distinction struck a powerful chord in the field, and Riskin’s “spectrum” of practice was soon translated into a typology of models of practice—the facilitative and evaluative models. According to one authoritative view of the facilitative model, the mediator in that model acts solely as facilitator or manager of the parties’ negotiation or problem-solving process.⁶³ The mediator establishes ground rules, facilitates information exchange, defines issues and structures an agenda, tries to generate movement toward agreement by various means (such as encouraging parties to focus on interests rather than positions, emphasizing areas of agreement, discouraging discussions of past incidents, and limiting expressions of intense negative emotions) and structures the closing of the discussions. However, the facilitative mediator expresses no views whatsoever on the merits of any substantive issue. By contrast, in the evaluative model, the mediator not only serves as process manager, but also offers expert case-evaluation (assessing strengths and weaknesses of each party’s case or proposals), substantive settlement recommendations (including, for example, predictions of court outcomes or other consequences), and strong pressures to accept those recommendations.⁶⁴ The articulation of the two approaches as distinct models of practice led to significant controversy—including argument over

⁶² See Riskin, *supra* note 57, at 111–12; see also Leonard L. Riskin, *Understanding Mediator Orientations, Strategies and Techniques: A Grid for the Perplexed*, 1 HARV. NEGOT. L. REV. 7, 24–32 (1996). Though Riskin actually offered a two dimensional “grid” including not only the facilitative/evaluative axis, but also a “broad”/“narrow” axis, it was the former that captured the attention of the field.

⁶³ See ALFINI ET AL., *supra* note 4, at 107–39; see also Nancy A. Welsh, *The Thinning Vision of Self-Determination in Court-Connected Mediation: The Inevitable Price of Institutionalization?*, 6 HARV. NEGOT. L. REV. 1, 15–21, 28–29 (2001). See generally Kimberlee K. Kovach & Lela P. Love, *Mapping Mediation: The Risks of Riskin’s Grid*, 3 HARV. NEGOT. L. REV. 71 (1998).

⁶⁴ See Riskin, *supra* note 57, at 111–12; Riskin, *supra* note 62, at 26–28.

whether an evaluative model could be called mediation at all.⁶⁵ Nevertheless, the distinction has persisted, and it has certainly undermined the homogeneity premise.

In the same year that Riskin introduced the distinction between facilitative and evaluative mediation, two other works appeared that also suggested distinctions between models of mediation. In a book that has become both influential and controversial, Professors Robert A. Baruch Bush and Joseph Folger identified and compared what they called the “problem-solving” and “transformative” models of mediation practice.⁶⁶ In the problem-solving model, as Bush and Folger describe it, the mediator focuses on achieving an agreement that solves the tangible problems involved in the parties’ dispute. More concretely, the mediator uses directive measures that control the process of discussion, so as to identify and narrow the problem, find a workable solution, and then persuade the parties to accept it in some form.⁶⁷ By contrast, in the transformative model, the mediator focuses on supporting each party’s development of clarity and confidence about their own views of the situation, together with understanding and empathy for the other party’s views—without any specific focus on resolution per se. More concretely, the mediator employs a variety of interventions to support party decisionmaking and interparty perspective-taking, but tries to avoid all forms of directiveness and to ensure party control over not only outcome but process decisions.⁶⁸ Like the facilitative/evaluative distinction, the transformative/problem-solving typology has been controversial,⁶⁹ but it has also gained currency and thereby weakened the premise that there is only one universal model of mediation in use.

⁶⁵ See Welsh, *supra* note 63, at 27–33; Chris Guthrie, *The Lawyer’s Philosophical Map and the Disputant’s Perceptual Map: Impediments to Facilitative Mediation and Lawyering*, 6 HARV. NEGOT. L. REV. 145, 146–54 (2001).

⁶⁶ BUSH & FOLGER, *supra* note 22, at 55–68.

⁶⁷ *Id.* at 63–71.

⁶⁸ *Id.* at 85–94, 100–01; see also Robert A. Baruch Bush & Sally Ganong Pope, *Changing the Quality of Conflict Interaction: The Principles and Practice of Transformative Mediation*, 3 PEPP. DISP. RESOL. L.J. 67, 85–96 (2002); Robert A. Baruch Bush, “What Do We Need a Mediator For?”: Mediation’s “Value-Added” for Negotiators, 12 OHIO ST. J. ON DISP. RESOL. 1, 29–32 (1996); Joseph P. Folger & Robert A. Baruch Bush, *Transformative Mediation and Third Party Intervention: Ten Hallmarks of a Transformative Approach to Practice*, 13 MEDIATION Q. 263, 267–74, 277–78 (1996).

⁶⁹ See, e.g., Carrie Menkel-Meadow, *The Many Ways of Mediation: The Transformation of Traditions, Ideologies, Paradigms, and Practices*, 11 NEGOT. J. 217 (1995); Neal Milner, *Mediation and Political Theory: A Critique of Bush and Folger*, 21 LAW & SOC. INQUIRY 737 (1996); Michael Williams, *Can’t I Get No Satisfaction? Thoughts on The Promise of Mediation*, 15 MEDIATION Q. 143 (1997).

Finally, in summarizing a unique study of the work of a dozen prominent mediators by an array of social scientists, Professors Deborah Kolb and Kenneth Kressel concluded that the practices of the mediators studied reflected two different models of mediation, which Kolb and Kressel called different “frames.”⁷⁰ They labeled these the settlement frame and the communication frame. In the former, the mediator focuses on producing a settlement, and does so by gathering information about the problem, developing an idea of what will solve it, and persuading the parties to accept some version of this solution.⁷¹ In the communication frame, by contrast, the mediator aims to “have the parties come away . . . with a different, better understanding of the problem, if not with a definite settlement,” and the role is to facilitate dialogue and enhance communication “in ways that further . . . better understanding and cooperation.”⁷² At one level, Kolb and Kressel’s models seem to parallel the problem-solving and transformative models identified by Bush and Folger. On closer examination, it could be argued that Kolb and Kressel’s models actually parallel the evaluative and facilitative models suggested by Riskin. However, whatever correspondence is seen among the three typologies, the Kolb/Kressel work reinforced the emerging recognition that mediation is not monolithic, and that there are indeed different models of practice in existence.

B. Implications: One-Model Tests and the Problem of Exclusion

In the field’s discourse, the idea of different models, which gained prominence beginning with the three works just discussed, has resulted in fairly broad acceptance of the view that there are three distinct models of mediation in use: evaluative, facilitative, and transformative.⁷³ For the present study, the important question is what impact this recognition of different models has had and should have on approaches to performance testing of mediators. In fact, that work has already had significant impact, both on the performance tests already analyzed in Part III above—referred to here, for convenience, as the TDP/progeny Tests—and on a number of other, “second-generation” performance tests developed more recently and discussed below.

⁷⁰ Kolb & Kressel, *supra* note 33, at 468–79.

⁷¹ *Id.* at 470–73.

⁷² *Id.* at 474–75.

⁷³ See, e.g., ALFINI ET AL., *supra* note 4, at 107, 140–47; Grace E. D’Alo, *Accountability In Special Education Mediation: Many a Slip ‘Twixt Vision and Practice?*, 8 HARV. NEGOT. L. REV. 201, 205–06 (2003); Neilson & English, *supra* note 8, at 223–24.

Given the features of the three models of mediation now recognized, it is clear that the preferred mediator behaviors that form common patterns in the TDP/progeny Tests closely match the facilitative model of mediation. As discussed in Part III, the mediator behaviors commonly preferred on those tests include: controlling and structuring the unfolding of the process, encouraging interest-based problem-solving, emphasizing common ground, focusing on the future rather than the past, and limiting expression of negative emotion.⁷⁴ These are the very elements of mediator practice seen as central in the facilitative model. In effect, the TDP/progeny Tests are geared to measuring mediator competence in the facilitative model, even if they may have been developed before the distinctions among models were fully articulated or widely accepted.

This correspondence between the TDP/progeny Tests and the facilitative model is even clearer when notice is taken of another mediator behavior preferred on at least some of those tests: refraining from giving evaluations or substantive advice. Thus, under "Ethical Behavior," the SDMC Test indicates that "the mediator [should] refrain from giving advice, opinions or judgments . . . or statements of preference regarding the law, facts or parties . . . [and] from offering advice or giving legal opinions. . . ."⁷⁵ Similarly, the Virginia and DON tests both measure whether the mediator "[a]voided giving advice, pressure, and judgment."⁷⁶ In short, these tests consider evaluative behavior an indicator of incompetence, because the model of mediation in which the mediator is presumed to be practicing is the facilitative model. Indeed, at least one of the tests explicitly states that "facilitative mediation is the model of practice."⁷⁷

Thus the recognition of distinct models of practice has had the impact of confirming the coherence of the TDP/progeny Tests and clarifying that they are good measures of competence within one particular model of practice—and good means of assuring that a different, nonpreferred model of practice is not being used.

However, this last point is actually quite a controversial one. It calls attention to the fact that, given the existence of distinct models of mediation, performance tests can function to approve practitioners of one model and simultaneously exclude practitioners of the others. In fact, that is what all of the TDP/progeny Tests do—some of them by their own admission.⁷⁸ The

⁷⁴ See *supra* text accompanying notes 58–59.

⁷⁵ SDMC Test, in *METHODOLOGY*, *supra* note 3, at 29.

⁷⁶ Virginia Test, *supra* note 48, § G; see also DON Test, *supra* note 49, § III.G.

⁷⁷ DON Test, *supra* note 49, § III (introduction).

⁷⁸ *Id.*; see also Spiegelman, *supra* note 7, at 707–08 (noting that the SDMC Test focused on facilitative mediation skills).

resulting rejection of transformative or evaluative mediators (and the implied judgment that they are “incompetent”) is certain to be controversial, especially where the test is used not as a means of quality control in a private, local mediation program, but as a “condition of admission” to a mediation roster in a statewide or other major mediation system, court-connected or otherwise.⁷⁹ This is the inevitable impact of continuing to operate on the premise of homogeneity in a world of practice that is admittedly heterogeneous or pluralistic.

The implication is that recognition of different models of practice should lead to a change in the way performance testing is conducted, a change that replaces the premise of homogeneity with one of pluralism. In fact, this type of change is already beginning to occur, and it is evident in what might be called a “second generation” of performance tests that represent a positive development in the direction of pluralistic performance testing.

C. Second-Generation “Multi-Model” Tests: Steps Toward Pluralism—and Problems

Each of the “second generation” tests discussed below is designed to accommodate more than one model of practice—in particular, the facilitative and transformative models—and therefore has a pluralistic character. In this sense, these tests represent progress toward a pluralistic approach to performance testing, which is the only sensible response to the recognition of multiple models of practice. However, as will be seen, the tests suffer from a common flaw, which limits their ability to function effectively as pluralistic quality assurance measures. The flaw is that these tests attempt to use a single instrument to measure competency in two (or more) distinct and different models of practice. The result of such “multi-model” tests is actually to reduce a tester’s ability to determine competence in one or both of the models included. Thus, while the development of these tests represents a step toward a pluralistic regime of performance testing, and is laudable for that reason, the step taken is itself problematic.

1. The Pennsylvania Test

The Pennsylvania Special Education Mediation Service (PaSEMS), established in 1988 to mediate disputes between school districts and families with handicapped children, developed a test to measure the quality of mediator performance in its program. The PaSEMS Test was developed in

⁷⁹ See *supra* text accompanying note 8; see also Neilson & English, *supra* note 8, at 227–29.

1999, based largely on the TDP Test and used the same general framework of categories and behaviors established in that test.⁸⁰ However, with awareness of the three different mediation models, and amidst differing views among the test's developers as to the value of each model in special education cases, the solution adopted was apparently to include measures of effectiveness related to all three models.

Thus, the test includes both a category captioned "Assist Parties in Generating Options" and also one captioned "Generate Options." Compare the high-score behaviors under each category:

- 7A. Assist Parties in Generating Options: Assisted the parties to develop their own solutions and to evaluate alternative solutions for themselves. Demonstrated commitment to allowing full play to parties' own values. . . .
- 7B. Generate Options: If and when mediator generated options directly, options were responsive to parties' concerns. . . .⁸¹

The test also includes both a category labeled "Assist Parties in Generating Agreements" and one called "Generate Agreements." Compare the high-score behaviors under each of these:

- 8A. Assist Parties in Generating Agreements: Emphasized areas of agreement. Clarified and framed points of agreement. . . . Packaged and linked issues to illustrate mutual gains from agreements. . . .
- 8B. Generate Agreement: . . . Effectively helped parties to move past apparent impasses. If substantive suggestions by the mediator were necessary, the suggestions demonstrated expertise . . . and were convincing.⁸²

Comparing the descriptions of high-score behaviors under each of these pairs of categories, it seems clear that 7A and 7B are inconsistent with each other, as are 8A and 8B. That is, the first in each pair calls for mediator assistance without advice-giving, while the second calls for direct mediator advice-giving. However, given the recognition of different models, the design of the test can be seen as an intentional attempt to include, and measure

⁸⁰ See D'Alo, *supra* note 73, at 224–28.

⁸¹ *Id.* app. B, at 261–62 (bullets added).

⁸² *Id.* app. B, at 263–64 (bullets added).

competence in, both facilitative mediation (7A and 8A) and evaluative mediation (7B and 8B).⁸³

Another example of “multi-model” testing in the PaSEMS Test is the inclusion of a category labeled “Understand Underlying Positions and Interests”—a category not found on any other test discussed in this Article. Under this category, high-score behaviors include:

Encouraged disputants to focus on concerns and interests. Demonstrated an in-depth understanding of . . . problems and interests not explicitly stated by parties. . . . Understood obvious aspects of the underlying reasons and interests of both sides. . . .⁸⁴

These indicators are clearly geared to assessing competence in facilitative, interest-based mediation. However, the language of 8B quoted above, preferring “substantive suggestions . . . [that] demonstrated expertise . . . and were convincing,” seems calculated to assess competence in evaluative, law-based mediation. Again, the explanation may be that the test is designed to assess competence in both models, or in either.⁸⁵

Nevertheless, the problem with such “multi-model” testing is that a mediator may score high in some categories and low in others precisely because she or he is practicing one model rather than the other. The result could be a poor overall score, though the mediator might be highly competent in one of the two models. The facilitative/evaluative tension in the above examples of competing measures of effectiveness is a good example of how this might occur. Good evaluative mediators who score well on categories 7B and 8B will score poorly on 7A and 8A, and the reverse will be true for good facilitative mediators. Simply put, it is difficult to measure competency in both approaches at the same time.

Another example of the test’s attempt to cover different models, and the inherent problem of doing so, is the inclusion of a category labeled “Move the Parties Toward an Improved Relationship.” In this category, high-score behaviors include:

Encouraged and facilitated constructive interactions directly between the parties. Established atmosphere in which anger and tension were expressed

⁸³ Alternatively, the test can be read as suggesting—especially in the “if and when” language of 7B and 8B—that effective mediation involves starting out in the facilitative model and then switching to evaluative practice if and when the situation warrants.

⁸⁴ See D’Alo, *supra* note 73, app. B, at 256.

⁸⁵ The alternative explanation offered above, that the test might actually favor a “combined model” mediator, seems less plausible here, since interest-based problem-solving and law-oriented evaluation are quite difficult to combine.

constructively. Emphasized areas of improved mutual understanding . . . helped improve the way parties viewed each other. . . .⁸⁶

These behaviors suggest use of the transformative model of mediation, with its emphasis on constructive interaction per se, rather than either of the other two models. However, the PaSEMS Test also includes a category—taken directly from the TDP Test—labeled “Manage the Interaction and Conclusion.” High-score behaviors in this category include “Made all decisions about managing the meeting . . . consistent with rationale for progress toward resolution. . . . Controlled process . . . ,”⁸⁷ clear indicators of competence in the facilitative model. Given the inconsistency between these two categories on the test (and as noted regarding the tension between facilitative and evaluative mediation), it is very possible that a good transformative mediator could score well on the “Relationship” category but poorly on the “Managing Interaction” category, with adverse impact on her overall performance assessment. The converse difficulty faces the good facilitative mediator.⁸⁸

In sum, a “multi-model” test like the PaSEMS Test is certainly a step away from the premise of homogeneity and toward pluralism. However, it carries inherent difficulties in application, precisely because it is difficult to measure competency in two (or more) models with the same test. Indeed, it may be that the kind of “single-model” tests discussed above in Part III are better at measuring performance than “multi-model” tests, because they limit their focus to one model and do not include inconsistent measures of effectiveness.

2. *The Family Mediation Canada Test*

Family Mediation Canada (FMC), a national organization of family mediation practitioners, adopted a performance test for family mediators in the late 1990s, as part of a mediator certification process.⁸⁹ The FMC Test was based on a test developed in the mid-1990s to evaluate the application of mediation skills among government-employed family mediators in British

⁸⁶ D’Alo, *supra* note 73, app. B, at 265.

⁸⁷ *Id.* app. B, at 266.

⁸⁸ Indeed, in a pilot evaluation using the PaSEMS test, the results showed that the majority of the mediators studied performed well on most categories, but poorly on both the “Relationship” and “Understand Interests” categories. *Id.* at 241–44. Perhaps the explanation is that those mediators were operating within a basically evaluative model, and therefore not employing facilitative *or* transformative practices.

⁸⁹ See Neilson & English, *supra* note 8, at 226–29.

Columbia.⁹⁰ According to an article by two of the test's developers, the FMC test was explicitly intended to accommodate "diversity of practice": "The certification process was designed to accommodate various . . . models of practice."⁹¹ Their account of the development of the test makes reference to both the facilitative and transformative models of mediation, and acknowledges that "researchers and mediators have identified and continue to analyze differences among mediators . . . and they continue to debate the correct theoretical model or models mediators should apply."⁹² The response of the FMC Test's developers was to try to design a "multi-model" test that would measure competency in more than a single model—a clearly pluralistic approach to performance testing. The following analysis examines this effort in some detail.

The FMC Test, without question, represents a serious effort to incorporate measures of competency in both transformative and facilitative mediation. Identifying the indicators on the test that relate to the facilitative model is not difficult, given the earlier discussion of the TDP/progeny Tests and the common measures of facilitative competency they utilize.⁹³ The same kinds of measures are recognizable on the FMC Test, as will be discussed. However, there was little discussion above of the kinds of behavioral measures that might indicate transformative mediation competency, other than a brief mention in the previous section analyzing the PaSEMS Test. Fortunately, a resource for identifying such measures is readily available, in this volume, in the article by Professors Della Noce and Antes and Judith Saul.⁹⁴ In that article, and based on research described there, the authors identify specific mediator behaviors characteristic of effective transformative mediation—which they call "supportive moves" (i.e., moves supportive of the general strategies required of a transformative mediator).⁹⁵ For the purposes of this Article, the behaviors described there—or similar behaviors—will be presumed to indicate competence in the transformative

⁹⁰ *Id.* at 227. The working group that developed the test included FMC, the Justice Institute of British Columbia, government-employed family mediators, and personnel and union representatives. *Id.*

⁹¹ *Id.* at 222.

⁹² *Id.* at 223.

⁹³ See *supra* text accompanying notes 35–59.

⁹⁴ Dorothy J. Della Noce et al., *Identifying Practice Competence in Transformative Mediators: An Interactive Rating Scale Assessment Model*, 19 OHIO ST. J. ON DISP. RESOL. 1005 (2004).

⁹⁵ *Id.* at 1049–53, app. A.

model.⁹⁶ Indeed, many such behaviors are included in the FMC Test, as will be seen.

The FMC Test measures specific behaviors in nine categories of activity, grouped into three general areas: managing the relationships, managing the process, and managing the content.⁹⁷ In each category, behaviors are identified as showing: no skills, inadequate skills, satisfactory skills, strong skills, or outstanding skills. From an overall examination of the test, it is clear that both transformative and facilitative behaviors are listed in almost every category. However, on closer examination a pattern emerges in which transformative behaviors often indicate low skills, while strong skills involve many more facilitative than transformative behaviors.

To aid in the discussion of these points, the following excerpt presents some of the specific behaviors listed under six of the test's nine categories⁹⁸ and "codes" the behaviors in terms of their connection to the two mediation models. Where behaviors are coded "F," the behavior is characteristic of the facilitative model, but *not* the transformative model. Conversely, where behaviors are coded "T," the behavior is characteristic of the transformative model, but *not* the facilitative model. Where behaviors have both codes, the behavior reflects both models.⁹⁹

⁹⁶ See Bush & Pope, *supra* note 68, at 85–96.

⁹⁷ See Neilson & English, *supra* note 8, app. A, at 237–44.

⁹⁸ *Id.* The Neilson/English version of the FMC Test is identical with the version currently in use, dated May, 2003. Phone interview with Carol McKnight, FMC Assessor (Oct. 13, 2003).

⁹⁹ To clarify and emphasize: the codes included in the following excerpt have been added by the author of this article, as discussed in the text above, and are *not* part of the text of the FMC Test itself. The published version of the FMC Test is itself coded to reflect what Neilson and English call "mediator behaviors associated with client transitions . . . and client self-determination. . . ." See Neilson & English, *supra* note 8, app. A, at 244. These codes are not reproduced in the excerpt here, since they were not designed to reflect, and do not necessarily reflect, the practice distinctions between facilitative and transformative models—the point of the analysis here.

FAMILY MEDIATION CANADA NATIONAL CERTIFICATION
PROGRAM

FAMILY MEDIATOR SKILLS ASSESSMENT CHECKLIST

I. MANAGING THE RELATIONSHIPS

...

I.B. Facilitates a collaborative relationship between the participants

...

2. Inadequate Skills

- Fails to establish effective guidelines (T)

...

- Does not mutualize (T)

3. Satisfactory Skills

...

- Promotes each participant's understanding of the other's point of view of the conflict (F,T)

4. Strong Skills

- Works with participants to promote mutual understanding, insight into, and empathy for the other (T)
- Helps participants find a mutual definition of the problem (F)

5. Outstanding skills

- Enhances participants' commitment to their new collaborative working relationship (F)

I.C. Manages power imbalances (passing score required)

1. No Skills

...

- Does not recognize unequal power balance (T)

2. Inadequate Skills

- Uses the participants' positional, emotionally laden, inflammatory language (T)
- Allows one participant to dominate the session (T)

...

3. Satisfactory Skills

...

- Ensures that both agree with process decisions (F,T)

...

5. Outstanding Skills

- Establishes and maintains a working environment based on equality (F)

II. MANAGING THE PROCESS

II.A. Attends to and explores participants' interests

1. No Skills

...

- Accepts statements at face value; no questions or probes (T)

...

2. Inadequate Skills

- ... [D]oes not explore interests before moving to solutions (T)

...

3. Satisfactory Skills

...

- Reframes positions and probes for underlying interests (F)
- Acknowledges the participants' feelings (F,T)

...

4. Strong Skills

- Identifies mutual and individual interests (F)
- Is able to differentiate and connect the interests of the participants (F)

5. Outstanding Skills

- Works with the participants to identify principles based on their underlying interests (F)

II.B. Manages conflict appropriately

...

2. Inadequate Skills

...

- Negatively reframes statements that serve to escalate, maintain, or entrench the participants' positions (T)

3. Satisfactory Skills

...

- Works with the parties to develop their communication guidelines (T)
- Establishes and maintains (redirects, refocuses) constructive negotiations (F)
- Establishes an emotionally and physically safe atmosphere (F)
- Ensures that focus is on the problem, not the people (F)
- Acknowledges and normalizes the participants' conflict (F)

4. Strong Skills

- Uses interventions to seek clarification (F)
- Confronts discrepancies (F)

...

- Reframes statements to defuse and gain consensus (F)

...

II.C. Evaluates ongoing process**1. No Skills**

...

- No control over process with no established and respected guidelines (T)

2. Inadequate Skills

...

- Allows participants to focus on past behaviors which are unrelated to the issue at hand (T)

3. Satisfactory Skills

...

- Maintains optimism and forward movement (F)

...

- Comfortable with silence (T)
- Helps generate an agenda and prioritizes it (F)
- Works with the participants to develop a process that respects their culture and their uniqueness (F,T)

4. Strong Skills

- Helps participants focus on issues to be addressed based on individual and mutual interests (F)
- Seeks clarification and direction from the participants in the process design . . . and makes procedural changes as necessary (F,T)
- Tracks body language . . . and attends to non-verbal cues (F,T)
- Encourages them in their capacity to work through the issues (F)

5. Outstanding Skills

- Consistently works with the participants to ensure . . . a mediation process . . . with predictable transitions and structure (F)

III. MANAGING THE CONTENT

...

III.B. Assists participants to apply interest-based solutions

...

2. Inadequate Skills

...

- Allows premature decisions based on insufficient information (T)

...

- Allows unrealistic and unworkable decisions (T)

3. Satisfactory Skills

...

- Accurately summarizes progress (F,T)
- Breaks solutions down into manageable portions (F)
- Promotes participants' ability to define their own outcomes or solutions (F,T)

...

4. Strong Skills

- Helps the participants identify principles and criteria that will guide their decision making (F,T)
- Helps participants select a wide variety of creative options that best address their mutual as well as individual interests (F)

5. Outstanding Skills

- Works with the participants to develop their own principles to evaluate their solutions (F,T)
- Encourages the participants' belief in their ability to use their own criteria to develop interest-based solutions for the present mediation and for their future negotiations (F)¹⁰⁰

The foregoing coded excerpt from the FMC Test is included here, despite its length, because this test is the most serious attempt thus far to accommodate different models of practice. It is therefore important, in any analysis, to give a full picture of the pattern of mediator assessment reflected in the test. Considering first the specific behaviors included, it is certainly true that both facilitative and transformative mediation behaviors are found on this test. In this connection, it is especially important to note the "caveat" that appears on the test itself: "You are not expected to demonstrate every one of these behaviors."¹⁰¹ This presumably means that a mediator can pass the test by demonstrating some, even if not all, of the behaviors that are satisfactory or better and avoiding most, if not all, of the behaviors that are inadequate or worse. Given this interpretation of the caveat, the inclusion of behaviors from both models should theoretically make it possible for a mediator to pass the test using either model. Moreover, behaviors from both models are, at least sometimes, rated as showing satisfactory or strong skills, which should strengthen the possibility of passage regardless of the model used. All this suggests that this "multi-model" test could indeed be effective in judging the performance of mediators using either model.

However, the matter is not so clear. A closer examination of the coded excerpt above shows two clear patterns that cast doubt on the proposition that the test can measure performance in both models equally well.

First, a pattern that is quite striking: *In all six categories excerpted above, transformative behaviors are rated as showing either "No skills" or "Inadequate Skills," while facilitative behaviors are never so rated.* That is, a mediator using the transformative model is at risk of receiving an unsatisfactory rating in *two-thirds* of the categories because, according to the

¹⁰⁰ See Neilson & English, *supra* note 8, app. A, at 237–44 ("F" and "T" codes added).

¹⁰¹ *Id.* app. A, at 237.

test, a mediator demonstrates unsatisfactory behavior if they *do not*: establish guidelines (ground rules) for the parties (I.B.2),¹⁰² “mutualize” (I.B.2), diagnose power imbalances (I.C.1), prevent “domination” by one party (I.C.2), ensure that parties articulate and explore “interests” (whether before or after discussing solutions) (II.A.2), probe for underlying interests (II.A.1), control the process (II.C.1), and ensure that decisions are not “premature,” “unrealistic,” “unworkable,” or based on insufficient information (III.B.2). However, in following the transformative model, a mediator will *intentionally avoid* doing all of these things, which are disfavored in the transformative model because they preempt rather than support party choice and agency in some significant way.¹⁰³ Nevertheless, in duly avoiding these behaviors, a competent transformative mediator will be at risk of an unsatisfactory rating on the test.

Furthermore, the transformative mediator can also expect a poor rating because, in following that model, the mediator will *intentionally*: use the parties’ “positional, emotionally laden, inflammatory language” (I.C.2), reflect “statements which serve to escalate, maintain, or entrench . . . positions” (II.B.2), and allow parties to focus on past conduct “unrelated to the issue at hand” and support them in doing so (II.C.2). These are behaviors favored in the transformative model, because they support party choice and create the possibility of more authentic interparty understanding—even if they may also intensify or sharpen conflict.¹⁰⁴ However, they are all behaviors rated as unsatisfactory on the test. In short, much of what a good transformative mediator is likely to do in a mediation session will involve behaviors described on this test as showing unsatisfactory skills. By contrast, facilitative behaviors are *never* rated as unsatisfactory on the test. The overall impact may be to put the transformative mediator at a comparative disadvantage, despite the intention to construct a “multi-model” test.

Second, at the other end of the rating spectrum, another pattern seen from the test excerpt is that, while transformative behaviors are sometimes rated as showing “strong” or “outstanding” skills, facilitative behaviors are rated in this way much more consistently. Indeed, in most of the categories, almost all of the behaviors rated above satisfactory are “F” behaviors (i.e., behaviors characteristic of the facilitative model but *not* the transformative model). The

¹⁰² The parenthetical references are to the relevant sections from the FMC Test excerpted in the text above.

¹⁰³ See Della Noce et al., *supra* note 94, at 1049–53, app. A; Bush & Pope, *supra* note 68, at 85–96.

¹⁰⁴ See Della Noce et al., *supra* note 94, at 1049–53, app. A; Bush & Pope, *supra* note 68, at 85–96.

logical impact of this is that, while a competent transformative mediator is likely to accumulate low ratings for many behaviors as just discussed, she or he is not so likely to receive high ratings from other behaviors to offset the lower ones and balance the picture. By contrast, the competent facilitative mediator is unlikely to receive low ratings in any category, and is likely to accumulate high ratings in several.

It is possible to argue that the potentially disparate impact just described is offset on the FMC Test by the way in which it is administered. First of all, assessors are themselves aware of the different models and can therefore discern when the mediator's behaviors form an overall pattern showing use of the transformative rather than the facilitative model; they can then score the test accordingly, adjusting what might otherwise be "unsatisfactory" ratings in light of the model being used.¹⁰⁵ Moreover, the test was designed to be "scored" by the assessor only after the mediator being tested has prepared a "self-assessment." That is, the mediator reviews his or her own performance (on videotape) with the test in hand, and has the opportunity to note behaviors that might be rated "unsatisfactory" and offer explanations and justifications, whether by reference to the situational context or—more important for the present analysis—by reference to his or her choice of mediation model, which the mediator is asked to describe.¹⁰⁶ In this way, a transformative mediator can give the assessor a basis for a positive interpretation of behavior that on its face would be rated "unsatisfactory," and can thereby pass the test. This kind of flexibility is, in short, built into the process in which the test instrument is applied. Indeed, in at least one other jurisdiction in which the FMC Test has been adopted, this flexible approach is also followed, and mediators are asked in advance to provide a "self assessment" statement describing their model of practice, which is then forwarded to the reviewer, who uses it to put the mediator's performance in context.¹⁰⁷

Still, the need to make adjustments in the administration of a "multi-model" test like FMC's suggests that the test as such does not fully do the job it is intended to do. Indeed, it may be that it is impossible for a "multi-model" test to do that job, because it is simply too much to ask of any single

¹⁰⁵ Phone interview with Carol McKnight, FMC Assessor (Oct. 13, 2003).

¹⁰⁶ See Neilson & English, *supra* note 8, at 231–32; email from Carole McKnight, FMC Assessor, Dec. 7, 2003.

¹⁰⁷ Phone interview with Robert Ketcham, Certification Committee Co-Chair, Maryland Council for Dispute Resolution (MCDR) (Oct. 20, 2003) (confirming MCDR's adoption of the FMC Test and explaining procedures for using it).

test.¹⁰⁸ This argument gains strength from the fact that the imbalance in the FMC Test lies not only in its details, analyzed above, but in a more general pattern of giving more regard to the facilitative model than the transformative model. That pattern can be seen in its very definition of the general activities or tasks of the mediator. Thus, five of the test's nine categories deal with activities that, in the transformative model, are simply not part of the mediator's job: managing power imbalances (I.C), attending to and exploring parties' interests (II.A), managing (i.e., controlling) conflict (II.B), controlling the process (II.C), and assisting parties to apply interest-based solutions (III.B). As discussed above, process control, power-balancing, and facilitation of interest-based bargaining are all central mediator tasks within the facilitative model.¹⁰⁹ Within the transformative model, however, they are not mediator tasks at all.¹¹⁰ Thus, over half of the FMC Test deals with activities that are simply not part of the transformative mediator's job, understood on its own terms.

Meanwhile, activities recognized as key parts of the job of a transformative mediator are included in the FMC Test only indirectly, and not as categories of their own. As described by Della Noce, Antes, and Saul, effective transformative mediation involves five major strategies, including orienting the parties to their own agency (power of choice), supporting their "conflict talk," and supporting their decisionmaking process.¹¹¹ In a test geared to measure competence in the transformative model, each of these would receive separate treatment as a category of mediator activity, with ratable behaviors, like the nine categories of the FMC Test. Indeed, the assessment model described by Della Noce et al. does precisely this.¹¹² On the FMC Test, by contrast, assessment of these three activities, to the extent

¹⁰⁸ It is probably also too much to ask of an assessor applying the test. Despite the possibility of an assessor's making adjustments in an applicant's score in light of the use of the transformative model (*see supra* text accompanying notes 104–106), it might be quite difficult in practice for the assessor objectively to score those adjustments. This is particularly so since some of the behaviors scored as "no skills" or "inadequate skills" on the test would show high skills in the transformative model. *See supra* text accompanying notes 101–103. Thus, the assessor might have to effectively reverse the scoring called for by the test in order to "adjust it" adequately for the transformative mediator. Placing this kind of burden on—and discretion in—the assessor might detract from the objectivity of the test and thereby undermine its legitimacy. *See, e.g.,* Stephanie A. Henning, *Note: A Framework for Developing Mediator Certification Programs*, 4 HARV. NEGOT. L. REV. 189, 224–25 (1999); Donald T. Weckstein, *supra* note 12, at 784–85.

¹⁰⁹ *See supra* text accompanying notes 35–59.

¹¹⁰ *See* Bush & Pope, *supra* note 68, at 77–85; Della Noce et al., *supra* note 94, at 1049–53, app. A.

¹¹¹ Della Noce et al., *supra* note 94, at 1025–36.

¹¹² *Id.* at 1022–46.

that it occurs, is diffused and treated only indirectly in some of the test's essentially facilitative activity categories. In short, the bulk of what a transformative mediator is supposed to do, as defined by the model itself, is not given the same kind of direct, focused attention on the FMC Test, as is given to the "core activities" of the facilitative mediator.

At an even broader level, the FMC Test pictures the mediator's job as "manager" of the relationships (I), process (II), and content (III) of the mediation—a vision consistent with the facilitative model. In the transformative model, by contrast, the mediator's job is "supporter" of the parties' efforts to change the quality of their interaction.¹¹³ A "managerial" mediator role, whether regarding process or content, simply has no place in this model. Yet the FMC Test adopts this terminology as its broadest level of analysis.

In short, even though the FMC Test was intentionally designed to accommodate more than a single model of mediation, it may be that it does a much better job of testing for competence in the facilitative model than any other. There is some suggestion, even in the account of the test's development, that it was still rooted in "single-model" thinking. Thus, the authors of that account, both participants in the test's development, explain that the test "respect[s] a diversity of theoretical and methodological approaches to mediation, *provided that* the approaches are consistent with the fundamental qualities of mediation identified by practitioners."¹¹⁴ Seemingly advertent to one of those "fundamental qualities," they state that "one of the key tasks in all mediations is the identification and articulation of disputant interests."¹¹⁵ However, as discussed above, there is a difference between followers of the transformative and facilitative models about whether "articulation of interests" is a key task or fundamental quality of mediation. Indeed, this is only one of many significant differences between the models, and those differences are the real meaning of "diversity of practice."

The FMC Test certainly represents a step forward in the direction of pluralism in performance testing—as does its adoption by the Maryland Council for Dispute Resolution (a statewide organization of Maryland ADR practitioners), which had previously used a test quite similar to the

¹¹³ See Bush & Pope, *supra* note 68, at 83–84.

¹¹⁴ Neilson & English, *supra* note 8, at 225 (emphasis added).

¹¹⁵ *Id.* at 224. Indeed, Neilson and English introduce their discussion with the statement that "we believe that a collaborative, facilitative and interest-based approach works, *not only within mediation*, but also in designing accreditation standards for mediators," a statement that strongly suggests the view that the facilitative model of practice is preferred. *Id.* at 222 (emphasis added).

TDP/progeny Tests.¹¹⁶ Open recognition of the different models of mediation, and sincere attempts to account for the differences between them in performance testing, is genuine progress. Nevertheless, the analysis offered here suggests that, while the FMC Test is an attempt to go beyond a single-model test of the TDP type, it has not gone far enough in fully recognizing and accommodating different models of practice. Indeed, the analysis suggests something more—that no single performance test can effectively or fairly accommodate mediators using entirely different models of practice. Trying to do so with a single test—even a supposedly multi-model test—presents a challenge, both to those who draft the test and to those who have to apply it, that probably cannot be coherently, fairly and effectively met. In order for performance testing to respond to the multi-model world of actual mediation practice, it must go beyond not just single-model thinking, but single-test thinking. Performance testing must move fully into a pluralistic mode.

V. LOOKING INTO THE FUTURE: PLURALISTIC PERFORMANCE TESTING—MULTIPLE TESTS FOR MULTIPLE MODELS

As noted much earlier in this Article, resistance to performance testing has been based in part on two reasons: the difficulty of gaining consensus on the set of skills a competent mediator should possess, and the concern that competent mediators would be excluded if skill sets for testing were constructed without a solid foundation.¹¹⁷ In fact, these reasons represent two sides of the same coin, the same flaw that underlies all efforts thus far to construct useful mediator performance tests. That flaw is the insistence on trying to address distinct and different models of practice in a single, comprehensive evaluation instrument. One major reason it has been so difficult to gain consensus on what skills a competent mediator should possess is that there is no such thing as mediator competency in a general sense. The meaning of competency differs in different models, and what is competent practice in one may well be incompetent practice in another.¹¹⁸

¹¹⁶ Phone interview with Robert Ketcham, Certification Committee Co-Chair, Maryland Council for Dispute Resolution (MCDR) (Oct. 20, 2003) (copy of previous test on file with author).

¹¹⁷ See *supra* text accompanying notes 6–9.

¹¹⁸ See Della Noce et al., *supra* note 94, at 1014–15; Dorothy J. Della Noce, *Ideologically Based Patterns in the Discourse of Mediators: A Comparison of Problem-Solving and Transformative Practice* (2002) (unpublished dissertation on file with author, presenting research documenting that “good practice” (as understood by mediators themselves) in facilitative model is converse of that in transformative model, and vice versa).

Discussions of competence without reference to specific models are therefore inevitably unproductive. Moreover, if somehow such discussions do lead to the definition of skill sets for testing, the resulting tests are inevitably likely to privilege one model of practice over others, despite the best intentions to avoid such disparity. That is the lesson of the second-generation, multi-model performance tests reviewed above.

A. Pluralism in Performance Testing: Looking Back and Looking Forward

Just as the field has accepted that there are different models of mediation, it is time to accept that, if performance tests are to be fair and useful, there must be different tests for the different models of practice. Ironically, to get some idea of what such a “future” would look like, a good starting point is found in the earliest performance testing work, discussed in Part II above.¹¹⁹ In its response to the critique generated by the original TDP Test, the Test Design Project reconvened and developed a new document, *A Methodology for Use in Selecting, Training and Evaluating Mediators*.¹²⁰ That document discussed explicitly the newly emerging recognition of different models of mediation, and then incorporated not only the original TDP Test but also “variants” that contain different measures “associated with different approaches to mediation, such as settlement-oriented and transformative, or evaluative and facilitative” and “based on different models of mediation.”¹²¹ Though the variant tests themselves are incomplete, and the measures used in them insufficiently developed to reflect accurately the associated models, the conception of the TDP in its *Methodology* is clearly supportive of the idea of using different tests to assess competency in different models of practice. The *Methodology* still maintains that it is possible to describe a “core of behaviors” common to many mediators, and states that the “common core approach . . . continues to be the primary mode of discussion here.”¹²² It nevertheless acknowledges—and tries to give examples of—the possibility of a taking a genuinely pluralistic approach to performance testing, rather than using single, multi-model tests.

A second, more forward-looking example of what pluralistic performance testing might look like, although perhaps not intended as such, is found in the model developed by Della Noce et al. for assessing

¹¹⁹ See *supra* text accompanying notes 18–35.

¹²⁰ See generally *METHODOLOGY*, *supra* note 3.

¹²¹ *Id.* at 20–21.

¹²² *Id.* at 6.

competence in the transformative model.¹²³ In presenting that model, Della Noce and her colleagues specify for every category of activity, not only mediator behaviors that form part of good transformative practice (“supportive moves”), but also behaviors that have no place in such practice (“nonsupportive moves”).¹²⁴ In the appendix to the study, these transformative and nontransformative behaviors are charted side by side.¹²⁵ Examining the text and charts, it is clear that the behaviors listed as nontransformative in each category/chart are indeed foreign to the transformative model—but they would almost certainly be viewed as appropriate, and indeed highly competent, in a facilitative model of practice. In effect, in the process of designing a performance test for transformative mediators, Della Noce et al. have also constructed a test for facilitative mediators, and they have placed the two tests side by side in a way that inadvertently offers a good example of what a pluralistic testing regime might look like.

In such a regime, different, model-specific tests would be made available, and mediators would choose which test they want to have applied to them. Mediators practicing exclusively within one model could take the test for that model. Mediators who practice in both models, depending on their clients’ wishes, could take both tests, and if successful would demonstrate that they are competent in both. Assessors would not be required to “adjust” the application of single test in a subjective way to accommodate for different models of practice, since the tests themselves would provide that accommodation. The challenge of getting practitioners to agree on the content of a single, universal skills set would be avoided, since practitioners of each model would find it far easier to agree on the skills involved in their model. Fears of exclusion by practitioners of “disfavored” models would be reduced, since no model would be disfavored by use of a single, universal test; rather, mediators using each model would have access to a credential appropriate to their work.

It should be noted that the discussion above has largely limited its attention to two mediation models, facilitative and transformative, since these are the two addressed most often in the tests analyzed here. However, a truly pluralistic testing regime cannot afford to ignore the third important model of mediation in use today, evaluative mediation. It goes without saying that evaluative mediation involves practices different from both of the other models, and that testing for competency in that model therefore would necessarily require different performance measures than those for either of

¹²³ See generally Della Noce et al., *supra* note 94.

¹²⁴ *Id.* at 1022–36.

¹²⁵ *Id.* at 1049, app. A.

the other models.¹²⁶ Although it cannot be addressed within the scope of this Article, development of those measures, and of valid tests for competency in evaluative practice, is as important to the future of performance testing as any of the other issues addressed in this Article.¹²⁷

B. *The Challenge of Diversity, the Strength of Diversity*

The history of efforts at mediator performance testing has demonstrated that neither single-model tests nor even multi-model tests can adequately measure mediator competence in a pluralistic world of practice. Those efforts have also begun to demonstrate, however, that performance testing can work in that pluralistic world, if the testing approach is itself a pluralistic, multi-test regime. There are even examples emerging of what a multi-test system might look like, and further work can surely refine and develop that work further. There are, in short, no insurmountable conceptual or practical barriers that prevent the field from meeting the challenge of performance testing, and thereby strengthening quality assurance in a very powerful way. Rather than shrinking from that challenge, it should be faced with increased vigor, building on the foundation of the work done and lessons learned thus far, as reviewed in this Article.

Indeed, the lessons learned from the performance testing enterprise can usefully be applied to other work directed to quality assurance. Specifically, where rules and regimes are established to evaluate mediator training curricula, grant “advanced status” to practitioners, or even rule on questions of professional ethics, those rules should not ignore the pluralism of mediation models and therefore practices.¹²⁸ In fact, though performance

¹²⁶ It should be clear from the discussion of the various tests in earlier sections that, almost without exception, none of the performance tests analyzed here includes what might be considered measures of competent evaluative practice. On the contrary, many of them explicitly disfavor such practices. *See supra* text accompanying notes 75–79. The only one that seems to try to include measures of evaluative competency, although with some potential for confusion, is the PaSEMS Test. *See supra* text accompanying notes 81–85.

¹²⁷ *See* Robert A. Baruch Bush, *Substituting Mediation for Arbitration: The Growing Market for Evaluative Mediation, and What it Means for the ADR Field*, 3 PEPP. DISP. RESOL. L.J. 111, 124–26 & nn. 39–44 (2002); Ellen A. Waldman, *The Challenge of Certification: How to Ensure Mediator Competence While Preserving Diversity*, 30 U.S.F. L. REV. 723, 724–28 (1996).

¹²⁸ *See, e.g.*, Dorothy J. Della Noce, Robert A. Baruch Bush & Joseph P. Folger, *Clarifying the Theoretical Underpinnings of Mediation: Implications for Practice and Policy*, 3 PEPP. DISP. RESOL. L.J. 39, 59–61 (2002). This is a significant current policy issue in the Association for Conflict Resolution (ACR), the largest national organization of ADR practitioners, including mediators. For example, in considering establishment of

testing is an important area, it is still not very widely used. By contrast, rules that grant or deny approval for trainers to provide mediator training, and rules that define the boundaries of ethical and unethical practice, are very widespread and have enormous impact on the kind of practice that mediators are free—or not free—to study and engage in. The field cannot afford *not* to accommodate—and thereby support and assure quality in—the different models of practice that are in actual use today.

The claim has been made that recognition of such differences will be divisive, will undermine the ability of the field as a whole to maintain a common identity, and will therefore compromise efforts to reach out to the public and to relate to those in other fields.¹²⁹ In fact, if recent social history has taught anything, it is that recognition and accommodation of diversity brings strength, not weakness. If the mediation field can recognize and accommodate its own diversity, it will find new strength to encounter other professions—and new strength to help individuals and groups deal in a constructive and positive way with the conflicts that their own difference and diversity inevitably engender.

an “advanced practitioner” membership category, a work group recently recommended that,

AMP criteria should neither exclude nor privilege one model of mediation practice over another, nor should it create or support particular or exclusionary markets for training and education to potential applicants. . . . [C]are should be given not to create additional requirements that restrict membership access to particular models of and approaches to mediation practice.

ADVANCED PRACTITIONER MEMBERSHIP WORK GROUP, RECOMMENDATIONS FOR ADVANCED MEDIATION PRACTITIONER (AMP) MEMBERSHIP STATUS 5 (June 2003) (on file with author). The ACR Task Force on Mediator Certification, in its report to the ACR Board, also affirmed the need to accommodate different models of practice. *See* ACR TASK FORCE, *supra* note 6. A third ACR study, on the subject of mediation practice and the unauthorized practice of law, drafted a report identifying “proper” and “improper” mediation practice in this context. Without explicitly identifying the practices discussed in terms of a specific model, the report nevertheless distinguished between facilitative and evaluative practices in defining proper and improper practice—although it also reported that the task force members were not able to reach consensus on whether the evaluative practices were proper or not. ACR TASK FORCE ON THE UNAUTHORIZED PRACTICE OF LAW, DRAFT REPORT 23–25 (Aug. 2002) (on file with author).

¹²⁹ *See, e.g.,* Neilson & English, *supra* note 8, at 223–24.